

Abstraksi

Data alamat merupakan bagian penting dalam suatu *record* yang berisi informasi mengenai data pribadi seseorang atau organisasi, oleh sebab itu data alamat harus tersedia dalam kondisi yang baik sebelum diproses oleh data mining atau disimpan ke dalam *data warehouse*. Kenyatannya, data alamat biasanya ada dalam bentuk yang tidak konsisten, masih dalam bentuk *free-form*. Untuk menghasilkan informasi yang akurat, data alamat dalam bentuk *free-form* akan lebih baik jika ada dalam bentuk yang konsisten atau memiliki format tertentu, sehingga sistem akan lebih mudah untuk mengolah data alamat menjadi sesuatu yang memiliki nilai lebih.

Dalam Tugas Akhir ini digunakan *Hidden Markov Model* (HMM) untuk membuat sistem yang dapat melakukan segmentasi alamat, sehingga data alamat yang tadinya ada dalam bentuk *free-form* akan memiliki bentuk yang konsisten. Jenis HMM yang digunakan adalah HMM *ergodic* karena jika diilustrasikan, pola alamat harus mengakomodasi kebutuhan setelah bergerak ke arah kanan, maka harus bisa kembali bergerak ke arah kiri. Sebelum data alamat masuk ke dalam HMM, dilakukan proses data *pre-processing* yang mencakup *data cleaning* dan *data tagging*. Untuk mendapatkan nilai probabilitas *state transition* dan *observation symbol* digunakan *supervised* dan *unsupervised learning*.

Dari hasil sejumlah observasi menggunakan model HMM dengan jumlah *state* 11 dan jumlah *observation symbol* 10, sistem menghasilkan akurasi terbaik 93.33% pada data *testing* dengan jumlah data latih 80 dan menggunakan *supervised learning*.

Kata Kunci : data alamat, segmentasi, *Hidden Markov Model*, *supervised learning*, *unsupervised learning*